COGNITIVE TASK ANALYSIS OF THE WARNING FORECASTER TASK

Final Report
Order No. RA1330-02-SE-0280

Prepared by:
Bianka B. Hahn
Erica Rall
David W. Klinger
Klein Associates Inc
1750 Commerce Center Blvd. North
Fairborn, Ohio 45324-6362
PH: (937) 873-8166
FX: (937) 873-8258

Prepared for:
The National Weather Service
Office of Climate, Water, and Weather Services
Norman, OK 73072

December 31, 2002
Revised January 8, 2003

Table of Contents

Introduction

Imagine an inexperienced warning forecaster, trained to use advanced algorithms as the primary basis for generating warning forecasts. This forecaster generally issues warnings every time the system provides a flashing alert (an inverted red triangle) as to possible severe weather, and thus does not really pay attention to whether a storm is developing in the middle of uninhabited grassland or near the center of a metropolitan area. This forecaster's warnings are based on the appearance of inverted red triangles on a display, which are driven by predetermined algorithms. The organization's measurement of this forecaster's performance is based solely on the numerical count of the storms he/she has missed over the last year, regardless of the storm's severity or amount of damage sustained in lives or property.

On the receiving end of this forecaster's warnings is a citizen of Oklahoma City (OKC). He has no faith in the warnings issued by the weather service as the myriad of inaccurate and unspecific warnings has dulled his ability and desire to notice. He can't make sense of how severe the threat is on a particular day because they all sound alike to him. In short, he mostly ignores warnings. On the day of an F4 supercell tornado, our citizen does not know to look for shelter and does not know that a storm heading towards downtown OKC is going to develop a severe tornado that will claim many lives. He goes about his business, unaware that a weather emergency is taking place.

Is this a hypothetical situation? Yes. Is this the future? Maybe. In the future our safety could be based on numerical calculations that bypass human expertise and rely on mathematical models to predict chaotic events. These models treat every developing storm like any other storm independent of how strong it might become or who might be in its path. Fortunately, this citizen was just an invented person. Similar weather events really did occur but an experienced warning forecaster was in charge that day and was able to distinguish this storm from all the others he had seen over the years. He realized that this storm was going to be larger, stronger, and potentially more fatal than what OKC had ever seen before. It was his experience and expertise that made the difference; he was able to recognize that the warning he was about to issue needed to convey the imminent danger of this storm to the public in order to get a fast response from those in its path. He issued a "Tornado Emergency" that alerted the media and resulted in extensive reports broadcasted over radio and television convincing the OKC population to take shelter.

What expertise did this forecaster apply to allow him to make the judgment that this would be "a day like no other"? How do other expert forecasters pluck potential storms out of the wall of data that is presented to a forecaster? How do these forecasters gather clues, piece together cues, and recognize patterns? In the following report, we discuss our initial attempts to answer these questions.

Klein Associates conducted a Cognitive Task Analysis (CTA) in order to begin to capture the cognitive expertise of warning forecasters. Our CTA methodology includes tools for knowledge elicitation, analysis, and representation. In applying our methodology, we are aiming at capturing the cognitive aspects of the warning forecaster's expertise as it pertains to making an assessment of the developing weather situation, conducting frequent analyses of the incoming data, using available technologies to keep the big picture, and deciding to issue a warning.

Capturing this expertise has wide-ranging implications. Training developed based on expert knowledge will provide less-experienced forecasters with the skills to identify less obvious tornado signatures and develop forecasts that allow forecasters to stay ahead of the storm. Technologies developed with the expert user in mind allow the technology to become a logical extension of the expert's cognitive processes, thereby decreasing mental and physical workloads and supporting the expert's decision-making process.

This report will address our knowledge elicitation and data analysis methods, as well as the research results of this preliminary effort and our recommendations for future research.

Research Method Section

Subject-Matter Experts (SMEs)

During a two and a half day visit to the National Weather Service (NWS) Forecast Office in Oklahoma we had the opportunity to talk to six meteorologists with experience in warning forecasting. These SMEs traveled to Norman, OK to meet us from NWS offices spread across the country, and they represented offices located in Alabama, Oklahoma, and Texas. In addition, we later had the opportunity to conduct a phone interview with one additional experienced SME from Missouri; this brought our total sample size to seven.

Six of our SMEs had between twelve and twenty years of experience working as meteorologists. They had advanced through various positions within the NWS and currently held positions as either Science Officers (n=4) or Meteorologists in Charge (n=2). Through the years many had had the opportunity to work at offices located in different regions of the nation for years at a time. This allowed them to gain a variety of experiences with varying weather patterns as determined by the geography of the region and their area of responsibility. We therefore had the opportunity to talk to five forecasters with extensive tornado experiences based on the supercell type and one forecaster who was very experienced with squall lines and bow echo tornadoes.

One of our interviewees was a journeywoman who had just been promoted to a forecaster three months before. At that point, she had only forecasted two severe weather days, although one of those days had been extremely active in terms of the number of warnings issued.

Interviewing Method: The Critical Decision Method

Our researchers interviewed each SME for 1.5 to 2 hours. The interviewing method we employed was the Critical Decision Method (CDM), an in-depth interview method that explores one specific incident that the SME has experienced (Hoffman, Crandall, & Shadbolt, 1998). The incident must be the expert's own (i.e., it must come from his or her own, lived experience). The detail and specificity of the cognitive events that CDM is designed to uncover require first-hand experience—the individual must have seen, heard, smelled, touched, processed, and reacted for him/herself. And it cannot be a generalized account. It is not enough for a person to say, "Usually when this happens you will know it because X." CDM requires a specific case of X and a description of the event from beginning to end. A generalized account leaves out too much of the information we require, whereas the specific episode carries context with it that reveals how particular aspects and events in the environment impel the decision maker to action.

A CDM interview requires an initial step, that of guiding the participant to recall and recount a relevant incident. What makes an incident relevant—the type of task or event—depends entirely on the focus of the study, and the particular goals of the Cognitive Task Analysis. Once this step (also called the first sweep) has been accomplished, the interviewer conducts three more information-gathering sweeps through the incident. These sweeps are: Timeline Verification and Decision Point Identification; Progressive Deepening; and "What-if" Queries.

First Sweep: Incident Identification and Selection

We helped our experts with the selection of the incident by asking for an event that challenged their expertise as severe-weather forecasters, a time when they were glad to be the one in charge. Once the SME identified a relevant incident, he/she was asked to recount the episode in its entirety. We asked him/her to walk us through the incident, often starting with either the beginning of their shift that day or when they left to go to work in the morning to allow us to follow the entire progression of the incident.

Once the expert completed his/her initial recounting of the incident, we retold the story ourselves. The SME was asked to attend to the details and sequence in order to correct any errors or gaps in our understanding of the incident. The SMEs typically offered a few corrections and additional clarifying details on specifics of the technology or data they used. This critical step allowed the SME and the interviewers to arrive at a shared overview of the incident.

Second Sweep: Timeline Verification and Decision Point Identification

In this phase of the interview, the expert was asked to go back through the incident account a second time to structure and organize the account into ordered segments. It was our goal in this sweep to gain a better understanding of the timeframe of the developing severe weather event as the forecaster went through his forecasting process. Some of the important information that the experts noticed happened outside of their office, often early in the day from their homes or even on their drive to work. This information seemed to set the stage for the additional information-gathering and decision making later in the incident. We therefore frequently asked the experts to start telling us about their day even before their shift began. The timelines are representative of how the experts' thoughts developed about the problem of the day as time went on and the steps he/she took to interrogate the storm.

Third Sweep: Progressive Deepening and the Story Behind the Story

During this sweep through the incident we led the expert back over each identified segment of the incident account, while employing probes designed to focus attention on particular cognitive aspects of the incident and solicit information about them. The probes are designed to progressively deepen understanding of the event, and to build a comprehensive, detailed and context specific account of the incident from the perspective of the decision maker.

The nature of the solicited information depends on the purpose of the study. In this case, we were looking for the presence or absence of salient cues in the environment, the specific nature of these cues, assessments of the situation and the basis of those assessments, expectations of how the situation might evolve, and options evaluated and chosen. We were also interested in the role and value of individual team members to the expert's decision-making process in the incident.

Fourth Sweep: "What-If" Queries

Due to some time constraints we did not have the chance to complete this sweep in every interview. This sweep in general is valuable in that it helps to uncover where the decision process might break down. It points out vulnerabilities of a process but also highlights where expertise is required. We used it in one instance to explore what it meant to be ahead of the storm and when this advantage can be robbed from the expert. It gave us some insight into the importance of functional mental models and the need to not purely rely on technology to make a warning decision.

Analysis and Representation

Analysis

The data analysis phase began with team brainstorming sessions to discuss cognitive themes and trends that we heard during the interviews. Then, we reviewed and coded the interview notes looking for those themes and trends. To accomplish this, we divided the interview notes so that each team member would conduct an in-depth review of two of the six interviews. Our strategy during this activity was to generate more questions for the particular SME regarding areas in which we felt there was more to be learned about their expertise, cues used, strategies, and decision process. This activity not only elicited a wealth of questions that could be pursued if the opportunity arose for future data collection, but also gave the team a better sense of the range of cognitive information available in the current data.

We also began working through each of the interviews to reframe the elicited incidents in terms of perceived cognitive phases. This process involved analyzing how the SME's mindset or cognitive approach evolved as the severe weather incidents progressed. For each of these phases we extracted pertinent cognitive information such as: which cues the forecaster noticed; how those cues were assessed; how the forecaster generated projections of how the situation might unfold; information he/she sought out; strategies for dealing with conflicting information or multiple storms, etc.

In the next step, we identified across all interviews the following areas where the forecasters' expertise mattered in the warning forecasting process: The forecasters' relationship to the public; implications of the forecasting team; their use of technology; their ability to use unusual cases to build up an experience base; their general approach to the weather; their mental model of severe storm development; and the expert decisions they have to make.

Representation

We developed a Cue Inventory (see Appendix A) to represent how the forecasters' expertise expresses itself in the wealth of information and meaning they are able to extract, often from single cues. The Cue Inventory is a 3-column table that identifies associated sources, cues, and assessments/projections/concerns. The source column includes where the expert received the cue, the cue column provides the cue or set of cues, while the assessment/projection/concern column describes the significance of that cue within the particular situation, what assessment or projection the expert was able to make based on the cue or set of cues. Occasionally cues also alert the expert to certain concerns and possible next steps that might have to be taken; these are also captured in the third column.

We assembled the Cue Inventory by first extracting information from each interview, which left us with a number of empty cells (see Appendix B). We then asked each SME (via email) to fill in the missing information for their incident account. This process not only supplied missing details for our inventory, but also allowed the SMEs to confirm our existing representation of their incident. Once we received the individual inventories from our SMEs, we developed a comprehensive inventory by collapsing the information across all interviews. We organized the inventory by the following categories: Significant Cues Noticed from Home; Significant Cues on the Drive to Work; Significant Cues at the Office. The choice of categories naturally fell out of the flow of the interviews. Within each of these categories we identified the source that provided the SME with the cue, the cue itself and finally the Assessment (A), Concern (C), Projection (P), or Course of Action (COA) that resulted from the recognition of the cue. It has to be noted that the assessments, projections, concerns, and courses of action developed from these cues were specific to the context of a particular situation. The order of the cues as presented here is not meant to be chronological but rather a representative sample of significant cues as identified by our SMEs.

Research Results

The purpose of this study was to use CTA methods to investigate the expertise of warning forecasters. It is important to note that as a short, preliminary study, our main focus was on identifying rich areas for further research to benefit the future development of training and technology design. That said, we were able to begin identifying characteristics of expertise in this domain, including: common attitudes/approaches towards the public, other forecasters on their team, technology, and the weather itself; types of decisions in which their expertise has the greatest impact; analysis of unusual cases to build up their experience base; and their assessment of severe storm development. In the subsequent sections of this report we will describe our preliminary findings for each of these areas as well as note rich cognitive areas of expertise that could be uncovered through further research.

Findings

Relationship with the Public

One aspect of expertise was the sensitivity the forecasters had to their role in protecting the public. They were very aware of storms' proximities to metropolitan areas, and this factor almost always played into their decisions of when and how to warn. They know when public weather awareness is more urgent, and when the public is not concerned "enough" in a particular situation.

Because of this concern for the public's safety, forecasters also take extraordinary measures to alert the public to coming dangers. For example, they might call ESPN event organizers when they know there is a major sporting event taking place in the path of a potential storm, or call the Storm Prediction Center (SPC) to request they upgrade their products to reflect the severe weather potential for that day. A number of forecasters mentioned using unusually ominous language in their warnings to get the attention of the media and the general public. These forecasters have developed sensitivities and strategies to get important weather messages out to the people who need to hear it, to get them to take precautions when necessary, and as a result, to save lives.

More extensive research could be done in this area on the effect of particular warning language on the public reaction. Several useful cues for taking the awareness pulse of the public and effective strategies for alerting the public in the course of a severe event were elicited in this study, but many more could probably be uncovered and documented through further targeted research.

The Team Aspect

Through the incidents we collected, it was apparent that the expert forecasters developed strategies to take advantage of interactions with other members of the forecasting team in their offices. The more experienced forecasters did not attempt to dominate difficult situations with their experience during severe weather incidents but rather leveraged opportunities presented *within* these incidents to interact, teach, and share insights with their colleagues.

In more than one incident we heard that an experienced forecaster took the event of the day as an opportunity to provide on-the-job-training to less-experienced colleagues. Some of those colleagues had either less experience as forecasters in general or were recently transferred from a different region of the country and had less experience with the weather patterns specific to that area. In these situations the expert communicated to the colleagues what he saw on the different screens, how he interpreted the weather patterns, and what he expected to see in the next data set.

In other cases we heard experts used their colleagues as sounding boards for developing their mental model of the storm. The experts sought the input and feedback of their colleagues to either share general opinions about their need for concern about the weather, to confirm suspicions, or to reevaluate their mental model of a developing storm. They valued these interactions as a crucial part of the forecasting process.

The forecasters all seemed to share a great appreciation of the need and importance of developing a shared situational awareness of the developing storm across the personnel in the office. They were strongly aware of the possibility of losing control over a situation if they became too focused on the smaller tasks at hand. In order to counteract this possibility many offices utilized the position of a coordinator during busy days. This person's responsibility is to keep the big picture in the forecasting area, communicate information to those who need it, find holes in the process, and find ways to fix those holes. This position becomes even more important when the workstations are distributed and lines of communication and awareness of each other's work are easily disrupted. The impact of the distribution of the workstations and personnel on the team's shared awareness, ability to anticipate, generate expectations, coordinate and synchronize must be lessened by this coordinator position.

We heard of numerous instances in which individual forecasters needed to compensate for other forecasters. In the heat of the situation, individuals are forced to focus in on certain details for a period of time while losing part of the big picture. When this happens, other forecasters must compensate in appropriate ways. For example, when one forecaster is writing a warning and digging down in the data to understand a particular storm, another forecaster needs to keep the big picture and watch for cues and patterns that might indicate a new storm or developing supercell.

Our data indicated that the more experienced forecasters were in tune with and aware of their colleagues' comfort and experience levels with either the use of specific technology or the forecasting of a certain weather event. Even in very busy situations many experts kept an awareness of the workload level of some of their less-experienced colleagues and, if necessary, were able to provide support.

Use of Technology

Every SME we interviewed believed that the mathematical-based prediction algorithms are "not the sacred truth" but are useful as safety nets or an attention management tools. Experts described situations in which they had to focus their attention very closely on the development of one particular storm with the highest probability to turn severe, while giving less time to other less probable or less severe storms. In these situations, algorithms, particularly the TVS (Tornado Vortex Signature), can be used to alert the forecaster to additional storms that might be increasing in strength.

In addition, it seemed the forecasters were well informed about the capabilities and limitations of the individual technologies and algorithms that were available for them to use. It was not uncommon for experts to cite the statistical failure rate of individual algorithms as it pertained to false alarm rates or oversights. The experts had certain technologies that they found to be more trustworthy than others; this assessment was often based on their ability to follow the "reasoning" of the algorithms. It appeared that few algorithms allowed the forecasters to drill down to the base data to confirm that the prediction was trustworthy. Over time, the forecasters have come to distrust a number of tools available to them because of their disability to confirm, trace, or understand the "reasoning" of the technology. But contrary to possible beliefs, expert forecasters do not mistrust and dislike all technologies. We found that many are enthusiastic about trying out new technologies and are hopeful for new technologies to become more user-friendly.

The experienced forecasters do not rely on or wait for technology to tell them when to issue a warning because they feel strongly about their ability to be ahead of a storm based on their decision-making processes. Experienced forecasters use a process (described in a later section) that allows them to feel confident about their decisions. They therefore do not allow either technology or colleagues to tell them when to make a warning. Our SMEs were very confident in their abilities to make sound and defendable decisions.

The most common desire for technology improvement stated by the SMEs was related to decreasing the delay of new data from the radar scans. The forecasters all wished the data scans would be more frequent than the current 5-minute delay. We understand that this is being addressed by current research goals within the weather service. We question the impact of this decrease in delay on the presentation of the data. If we can decrease the delay, can we then generate better interface solutions to present the data to forecasters in a way that supports their decision process? The SMEs, particularly those outside of OKC, would like to generate a better network of observers. Some parts of the nation are not as densely covered with radar sites and ground observers, making forecasting more of a challenge as these forecasts must be based on less complete information.

Experts agreed that they are not really looking for *more* algorithms that are designed to replace the human in the loop as the decision maker, but would rather focus future efforts on developing better-designed ones that make sense to the decision maker and can therefore be trusted.

Using Unusual Cases to Build Up Experience Base

Experts recognize the importance of building lessons learned. They understand the importance of seeing first hand the damage after a storm and getting feedback on fatalities, injuries, property damage, transportation effects, etc., and then relating these effects to the signs they noticed during the storm's development. A number of the forecasters we interviewed applied knowledge learned from this type of previous research and/or investigations in the critical incidents they related to us.

For example, one forecaster was able to instantly recognize a subtle signature in the velocity data, recognize that the storm was directly over the radar, and immediately issue a warning for the metropolitan area. He was able to make this invaluable, nearly instantaneous, connection because the pattern he saw on the radar screen matched a pattern he'd seen in an investigation years before. In that previous case, he and other forecasters had poured over the data for hours, finding only this one persistent signature to explain the phenomenon.

Building up a repertoire of these unusual cases allows forecasters to more efficiently diagnose the situations they are faced with in real life. Instead of spending valuable time connecting the factors and projecting the causes and effects, they can instantly match the patterns they are seeing to patterns in the current data. The larger the repertoire of these cases, the quicker they can recognize the unusual situations that may face them (Klein, 1993).

General Approach to the Weather

The experts we talked to in this domain had a unique, dynamic approach to the weather in general, which in turn affected their handling of severe weather situations.

First of all, these expert forecasters “live” the weather. They do not limit their assessment of the weather to their workstation at the NWS office. They check information on the Internet from home, they smell the moisture of the air when they walk outside their front doors each day, and they assess the movement of the clouds on their rides to work.

In a previous CTA study of weather forecasters (Pliske, Klinger, Hutton, Crandall, Knight, & Klein, 1997) the researchers broke the sample into categories, or types of forecasters. The most experienced forecasters were what the researchers called the “intuitive scientists,” and this is what we saw in our experts. According to Pliske et al. (1997) this group of forecasters seems to love the weather at some basic affective level. They treat the weather as a highly dynamic, “living” system that they are constantly striving to understand more completely.

We were somewhat surprised in our study to discover that *most* of the expert forecasters marked the beginning of their interest in meteorology with an early, memorable experience with severe weather. These vivid experiences as young children seem to have sparked in them a life-long fascination with the intricacies of weather. As further evidence of this fascination, we also noticed that many of the forecasters we talked to were avid storm chasers, at least at some point in their lives. In other words, these expert forecasters do not treat the weather as a 9-5 job, but as a life-long quest to understand the complexities occurring in the air above them.

This dynamic approach to the weather is important because it enhances their expertise as warning forecasters. Not only do they have the motivation to develop their understanding of the complexities of severe weather, but they also are able to incorporate cues and knowledge from their experiences outside work. They are able to add the full power of their senses (smell, touch, sight, etc.) to their performance on the job. We were surprised by the importance of some of the cues these forecasters noticed from home, or on their drive to work in the morning. These cues were especially important in getting a general sense of the instability in the environment. For example, one forecaster began suspecting a severe weather day from the smell of the air when he walked out of his front door that morning.

These expert forecasters are highly sensitive and constantly attuned to the potential for severe weather in the environment. Days ahead of time they are sometimes able to recognize that severe weather is on its way. This ability stems not just from their knowledge of weather patterns, but also from an active mindset; they are constantly assessing this potential in the environment, especially in severe weather-prone seasons of the year. They start playing scenarios in their head, hypothesizing about what might develop if certain cues line up, where might storms develop if the front goes the direction they expect, etc. This preparation allows them to anticipate when dangerous weather is most likely to occur, and allows them to prepare themselves and other members of their staff to handle the situation most effectively when it does occur. One forecaster we talked to was even able to plan staffing days in advance to accommodate handling the severe weather event on the day of its occurrence. They had the right team in place at the right time, instead of having to make the decision that something is happening, then call people into the office, and get them up to speed on the situation. This fully engaged mindset and senses of the expert forecaster allows them to anticipate and prepare for the arrival of severe weather.

Mental Model of Severe Storm Development

As mentioned in the previous section, warning forecasters often come into the office at the start of their shift after having collected some overall environmental assessments on their way to work, or by checking maps on the Internet. Once they arrive in the office they have a unique way of briefing themselves. This is a way to come up-to-speed quickly before they start talking to the forecasters on the previous shift about their assessments and predictions. At their desks, the forecasters configure their individual workstations in a way that suits their assessment style the best. The arrangement of the screen real estate seems to be a personal choice based on the season and the associated potential for severe weather. One strategy that an expert described to us for an uncertain weather day was to keep a variety of data sets available, in order to retain flexibility in making an assessment.

Experts "interrogate" the storms as they develop. They go through a process of proactive investigation of the characteristics and specifics of the developing storm. They choose different types of data sets on various screens to get a different view of the storm, and overlay different data sets to identify certain patterns. Through this process they develop certain questions they want the data to answer for them. Some of the experts run large sets of data past their eyes and cycle through different systems almost continuously, never sticking too long with any one data set in order to develop a better idea of the storm's development. In this process, the individual expert uses a large variety of different products to make his/her assessment.

Experts spend a lot of time projecting what will happen in the near future with a storm. This type of thinking allows the forecaster to stay ahead of the storm and not be reactive to already existing situations. Experts hypothesize about the worst-case scenario and play it out in their minds. They develop expectations about what they might have to see in the data for this scenario to play out. Each radar scan is used to compare the experts' mental models to the newly collected data to see if the expectations have been confirmed. The experts hypothesize that if a certain pattern in the next scan appears, then the storm is developing in a certain direction. In between the scans, the expert analyzes interactions between cues to answer questions like "does the reflectivity support the velocity couplet?" because if it does, a warning might have to be issued.

Over years of working with weather patterns our experts have developed "sensitivity for severity." What we mean by this is the ability to tell that a certain weather event is going to be much larger than what the forecaster experienced or saw before. In our interviews we heard reports about a number of career-changing events. Often they were able to realize at some point in the forecasting process that this storm will be larger, more destructive or faster developing than the usual storms that developed in the given region during that time of year. It was not unusual for our expert forecasters to be the first ones in their office to notice this development. But beyond the recognition of the possible magnitude of an event, they were also able to identify very early on which one storm out of a line of storms would develop to be "the one" to watch.

Without exception, all of the interviewed experts described the importance of seeking out 'ground truth'. In each of the events we heard about in our interviews the informative element of ground truth played an important role in the decision-making process. In some cases, it confirmed predictions about the storm's developments and helped in the follow-on warning decisions. In other cases it confirmed that a warning that was sent out was the correct choice for the specific region and provided post facto confirmation of a correct warning decision. Experts use all possible means to receive ground truth beyond the HAM radio reports and phone calls that often flood the office during severe weather. We heard forecasters report that they turned on the television and caught an ESPN report about the cancellation of a sports event based on severe rainfall, which in turn was a valuable cue that the weather system was carrying as much water as suspected. Other times, the cues were more obvious because the TV camera on top of a city building happened to capture the large funnel of a tornado that was moving towards the downtown. Experts take a moment to make phone calls to other offices to elicit possible ground truth reports that might not have been originally reported. They even might call the cell phone of an off duty forecaster in the hope that the forecaster is somehow out there looking at the storm. Experts seek out this type of information to cut down on uncertainty and remove the middleman, technology. The radar technology provides experts with indications of developing weather while ground truth often can further confirm the actual development.

Experts are skeptics and maintain their skepticism about any given weather situation for long periods of time. They are not easily swayed by convenient patterns in the data that might indicate one weather pattern over another. Experts are often not swayed by the opinions of their colleagues unless they find confirmative data that seems to exclude other options. We heard in more than one case that the warning forecaster in charge was much more skeptical about the developing weather situation than other people in their own office, and that it was his/her skepticism that, for example, allowed the forecaster to recognize in time that the developing weather system was not the typical tornado producer but a developing flash flood situation.

Experts generate a dynamic mental model of severe weather situations. After a few cycles of looking through radar data-expert warning forecasters develop a mental model of what they believe the incoming weather might be. They start to develop predictions or hypotheses of what the next data scan needs to show in order to confirm their mental models. They look through many different data sets, and data overlays, looking for certain cues in the data to confirm their model. Experts use their mental models to stay ahead of the storm and to be able to make sound and defendable decisions on where and when warnings should be issued. These mental models incorporate a strong understanding of weather patterns based on region and time of year. Our expert pool included experts from across the nation and their mental models varied quite drastically based on regional and seasonal differences. An expert from one region with sophisticated skills of developing and using his/her mental models will not be equally as efficient and successful in applying those if transferred to another region. Experts use their research experience as a basis for the development of their mental models. The knowledge they have gained through either investigating former severe weather cases or involvement in research on certain weather patterns allow them to build a framework that supports their mental models. In one instance, an expert's mental model allowed him to recognize certain weather patterns, even though he was working for the first time with the newly installed Doppler Rader. We saw experts that had a distinct number of available mental models based on former research of, for example, developing squall lines, and they were able to leverage those during their forecasting process. The preexisting mental models do not lock experts into a rigid framework of mind but rather provide a general guideline that allows for taking in new and surprising information, such as the much more rapid development of aspects of a storm, to adapt the existing model. Experts are able to learn from the surprises they encounter and use them as additions and refinements of their existing mental models to support future forecasting decisions.

Expert Decisions

We identified three types of decisions on which expertise has the most impact in this domain: Each decision will be explained in subsequent sections.

While expertise can have a subtle positive influence throughout an incident, certain decisions stood out as the high payoff decisions. We assessed expertise as having the most value in these situations for two reasons: 1) the difficulty of the decisions demands having expertise to make it, and 2) the impact of these decisions on the public has potentially huge consequences. We identified the following three decisions as fitting these distinctions; each decision will be explained in subsequent sections:

1. **Identifying "The Big One"**
2. **Predicting which part of a squall line will accelerate, or how much of velocity/reflectivity is enough to warrant the generation of a warning**
3. **Recognizing storm's severity, and the need to make this explicit to the public**

1. Identifying "The Big One"

One of the high-payoff areas for expertise in this domain is identifying "The Big One." By this we mean recognizing the potential for severity, identifying the one storm that distinguishes itself from other severe storms, in that it will cause the most damage, and that poses the most risk to the public. During seasons that are most prone to severe weather any given weather system usually does not only produce one storm but often dozens within rather short amounts of time. We heard about a day when 38 tornadoes developed in a few-hour period. Experts are often able to assess the developing weather system and pick out the one storm that will be most threatening. There are many factors that play into that assessment, including the speed of development of the storm, the direction it is moving into (metropolitan area vs. empty farmland), and the potential severity of the storm. Experts can assess these and other cues fairly early on and make a judgment to which storm to attend to first. This is not an easy decision because there are often fine nuances that the expert has to pick up to predict for example, if this storm is likely to shift in direction to head towards a highly populated area. Recognizing these nuances is an ability anchored in the experts' experience base and years of watching weather move across the region. They are able to pick up on early patterns of cues that do not mean much yet to a novice, but guide the decision process of an expert. The cost of missing "The Big One" can be excruciating for the public. Experts have therefore developed processes that allow them to stay ahead of the storm and make early warning decisions. The way by which they accomplish this is the leveraging of their mental models, their expertise, and a carefully chosen back-up system of technology as a safety net.

2. Predicting which part of a squall line will accelerate, or how much of velocity/reflectivity is enough to warrant the generation of a warning

This area of expertise is specific to the forecasting process that individual forecasters go through while evaluating the weather of the day to see the potential for severe weather. The process often starts when they get up in the morning and look out the window or check the Internet, and it continues throughout the shift at work when they realize which part of the weather front or which cell needs special attention. The expertise plays in specifically when the forecaster looks at the squall line and is able, based on his experience and mental model, to make specific predictions as to which part of the line will accelerate and where it will turn and cause damage. It is the fine set of cues that speak to the large amount of experience of the expert that allows the forecaster to make a warning decision for a specific area before the actual event is happening. We heard the same from forecasters whose expertise lies in the prediction of supercell tornadoes. Often they are faced with signatures on their radar that do not clearly indicate a tornado, yet they are able to make an accurate prediction of whether the storm will produce a tornado or not. In our interview with one less-experienced forecaster, we heard how hard it is to know how much velocity and reflectivity is enough to warrant the issuance of a warning on any given day. There are no exact guidelines, because at some times the reflectivity is strong while the velocity is supposedly not developed enough to indicate a tornado, yet the tornado develops. Experts have developed mental models that allow them to play out the individual scenarios and recognize patterns that novices are often not sensitized to. It is this expertise of knowing when severe is severe enough, recognizing the nuances of the storm's development even in face of irregularities that allows experts to make warning decisions to keep the public informed ahead of the storm.

3. Recognizing storm's severity, and the need to make this explicit to the public.

The recognition that the wording of a standard warning message will not be enough is our third high-payoff area of expertise. This decision encompasses a number of crucial assessments and an overall awareness of the public's "state of mind." The expert has realized based on the forecasting process he/she has gone through and the data he/she has looked and compared to his/her mental model that the storm is large and destructive in nature and, in addition, moving towards a populated area. The combination of assessments sets off an alarm in the forecaster's mind. The warning needs to go out but the question arises: is the public going to recognize the intense urgency and intrinsic danger of this storm? Experts often are aware of large outside festivals or sporting events in the area that might attract a lot of people and take this into consideration. When issuing a warning they also consider the media's current focus of attention on a sporting event or cues such as no mentioning of an earlier issued severe weather watch. In extreme situations, experts know to consider what it might take to get the public to become aware of the severity of the event, and they choose effective means. The wording of the warning message can catch the media's attention and reach a larger number of people.

Future Research Recommendations

To some, the future of severe weather forecasting is much like the future of air traffic control. That is, what role will technology play? There are those who believe that weather predictions and commercial airline traffic control can be done by machines. The tasks are so rule-based that human expertise can be captured and applied in a way that is superior to human performance. The thought, then, is that context-specific expertise is not required. At the beginning of this report, we presented a real-life scenario using a hypothetical citizen of Oklahoma City. Imagine for a second that you are that citizen and that the weather warnings you hear are all the same, are not always reliable, and are not overly accurate. Now imagine that you are in the sky over Texas and your aircraft is having mechanical problems. Would you want a machine directing you to an alternate landing strip? No. In both cases you want an expert in the field to provide you with the contextual information you need. The attempt to over-apply technology to fields in which detailed, context-specific expertise is required is simply wrong.

At the outset of this project, we were tasked with identifying expertise in the severe weather field and identifying potential areas where the entire system could be improved. With that focus, we have identified three areas in which we believe the National Weather Service can upgrade their products by applying what we know about expertise and decision making within the Weather Forecast Office.

> Area 1: Technology. Technology continues to push data collection and analysis in nearly every scientific field. The forecasting of severe weather is no different. So, what will the future of severe weather forecasting look like? Will we rely on a few machines that recognize dramatic data cues to issue watches and warnings? Is it simple enough that a machine can run endless pattern recognition algorithms? Can a machine pull out the exact right set of data from a crowded data stream and decide how that might influence other seemingly distant and unrelated data? We don't think so. But we heard that at least a portion of this is being considered.
>
> What is needed is a solid foundation upon which to build future technology. Simply providing more tools and faster processors is not the answer. The answer isn't even providing more rapid radar updates. The technology that is introduced must synchronize with the decision-making process of the forecasters. System designers must understand how the forecasters look for data, what data are important and what data are noise, when in the forecasting process the forecasters need certain pieces of information and when they don't, how they use the algorithms and when they simply turn them off, etc. It is the drilling down into the cognitive processes of the end-user that provides the needed foundation for technology development. Without this step, money is wasted and performance decreases.

Area 2: Machines simply cannot make accurate enough forecasts. These machines, when using the right models and the right data, can make global predictions that have the right amount of detail so that subtle errors have little or no impact. But, pinpoint accuracy in severe weather predictions is critical and those predictions can only be made by human experts. Therefore, a training package that trains the critical skills necessary for expert performance is of the utmost importance. It is not enough to teach people about the weather and how it works, what is required is a comprehensive study of the expertise of severe weather forecasting. Just what is it that makes some people better at it than others? It may be the case that a person who has been forecasting the weather for 25 years is simply not as good at it as someone who has been doing it for 10. So, experience is not the only factor. We need to understand those factors and apply them to a training program.

Area 3: Teamwork and team functioning within the weather station. We study teams and how they work. We have been involved in numerous design efforts of command posts, decision centers, nuclear control rooms, cockpits, and information centers. We find that it is often a difficult balance to design one of these decision centers (which we believe the forecasting center to be) with both steady state and crisis operations in mind. What is good for one is not always good for the other. We have found that designing a decision center requires an understanding of the decisions, the context surrounding those decisions, the information flow (this includes receiving the right information in the right format at the right time), and the coordination with other team members (both internal and external). We also understand that technology is one of those team members. As decisions regarding technology placement are made, one must consider the impact that it has on team performance. The ultimate goal would be to best utilize the space limitations, to minimize scramble time, to streamline data fusion, to provide the experts with what they need when they need it, and to support and promote decision making so that products are delivered to the public as efficiently as possible.

References

Hoffman, R. R., Crandall, B. W., & Shadbolt, N. R. (1998). Use of the critical decision method to elicit expert knowledge: A case study in cognitive task analysis methodology. *Human Factors, 40*(2), 254-276.

Klein, G. A. (1993). A recognition-primed decision (RPD) model of rapid decision making. In G. A. Klein, J. Orasanu, R. Calderwood, & C. E. Zsambok (Eds.), *Decision making in action: Models and methods* (pp. 138-147). Norwood, NJ: Ablex.

Pliske, R. M., Klinger, D., Hutton, R., Crandall, B., Knight, B., & Klein, G. (1997). *Understanding skilled weather forecasting: Implications for training and the design of forecasting tools* (Technical Report No. AL/HR-CR-1997-0003 for the Air Force Material Command, Armstrong Laboratory, Human Resources Directorate Brooks AFB, TX). Fairborn, OH: Klein Associates Inc.

Appendix A – Cue Inventory

SIGNIFICANT CUES NOTICED FROM HOME

Source	Cue	Assessment (A)/ Concern (C)/ Projection (P)/ Course of Action (COA)
Numerical model and satellite data on the Internet	Upper level disturbance moving through early in the day	A: If it gets through "early enough", might have enough destabilization for severe weather later in the day
Own analysis; Storm Prediction Center outlooks	Unseasonably strong weather system approaching	P: That day would be a bad event day
Own sense of smell (outside)	Smelled the Gulf of Mexico moisture in the air - humid, salty, similar to being near an ocean	A: Abundant moisture to fuel storms and increase rainfall efficiency of those storms C: Started thinking about flood threat
Visual cue outside	Noticed that the speed of the clouds was "different"	A: Fast-moving clouds indicated strong "low-level jet stream" transporting moisture/instability north
Visual cue outside	Severity of downpour	A: Atmosphere capable of producing huge amounts of water
Internet	Tornado watch issued for an area located west of County Warning Area (CWA), but moving east	A: Storms would impact CWA soon, and on top of earlier very heavy rainfall C: Additional flood threat
Internet radar	Strong vertical wind shear	A: Instability in atmosphere P: Potential for serious storms
Internet radar	Lack of supercells developing	A: Tornadoes not the threat that day

SIGNIFICANT CUES ON THE DRIVE TO WORK

Source	Cue	Assessment (A)/ Concern (C)/ Projection (P)/ Course of Action (COA)
Visual cue	Cumulus layer of clouds racing north	A: Air mass is destabilizing COA: Need to monitor the satellite imagery when I get to work
Tactile cue	Relatively cool morning	P: Need the sunshine to heat things up before storms will form
Radio and HAM radio (in car)	Nobody talking about the weather	C: Severe weather might surprise these people because they're not paying attention to it

Source	Cue	Assessment (A)/ Concern (C)/ Projection (P)/ Course of Action (COA)
Visual cue	Not much storm activity in the area	P: Instability would not be released until afternoon
Visual cue	Signs of low-level moisture	P: Instability would be large and supportive of severe storms in the afternoon
Visual cue	Wisps of gulf stratus clouds	P: Instability would be large and supportive of severe storms in the afternoon

SIGNIFICANT CUES AT THE OFFICE

Source	Cue	Assessment (A)/ Concern (C)/ Projection (P)/ Course of Action (COA)
Walking into the office (visual cues)	Casual conversations going on, no one paying attention to the developing cumulus towers on the satellite imagery, no mention of storms in the forecast at all	A: Poor situational awareness COA: Need to get people focused on the threat, drop subtle cues at first then become more forceful
People on previous shifts	No thunderstorm forecasted in the area despite the developing cumulus field near the metro area	C: Concerned about tornado threat
Satellite imagery	Seeing cumulus that were beginning to build or tower	A: Indication that the cap was weak C: “Red flag” COA: Watch each image as it came in to monitor the vertical development P: Would the cumulus break the cap?
Numerical data	Parameters were pretty volatile; substantial overlap of both strong instability and vertical wind shear parameters	A: Very conducive for storm formation; instability and shear parameters right for a severe storm, atmosphere a classic supercell developer
Doppler data	Reflectivity and velocity structure	A: Storm develops “nice reflectivity,” but velocity structure does not look good for tornado formation; environment not “great” for developing tornadoes

Source	Cue	Assessment (A)/ Concern (C)/ Projection (P)/ Course of Action (COA)
HAM radio reports	Some people seeing strong rotation; some seeing no rotation	A: Conflicting information COA: Rely most heavily on what off duty NWS spotter was seeing and what the radar was showing
Another forecaster calling in from another office	Report that looks like a tornado situation	COA: Looks at the same thing to assess A: The rotation does not look good down low; don't feel it is time to pull the trigger yet
Call from boss (out of town at a meeting)	Told to issue a warning	C: Doesn't feel ready to issue warning yet
Numerical data, radar, overall environment, cap	Reflectivity, velocity signature	A: Storm had very good reflectivity, very good velocity signature, overall environment favorable for severe storms, including no strong cap P: This storm would produce a family of tornadoes
Call from off-duty forecaster chasing the storm; radar	Forecaster's description lines up perfectly with radar information	A: Convinced that warning needs to be issued immediately
Doppler Radar, knowledge of river basin layout in metro area, knowledge of past flood events	Storm moving slowly along the river basin	C: Storm could produce huge amounts of water A: There's an increased chance for flash floods
Doppler Radar "loop" of storm movement	Reflectivity	P: Storms moving only slowly east will cause increasing rainfall rates; merging storm cells will also increase rainfall rates
Doppler Radar 4-panel display of storm structure	Vertical structure	A: Generally "low echo centroid" cells suggesting very heavy rainfall
Doppler Radar "loop" showing storm movement	Line of thunderstorms approaching metro area	A: Not strong enough to be severe; past reports and radar indication suggested not severe or only marginally severe P: Wonders if they will be severe in the near future
Doppler Radar reflectivity image	Hailstorm moving towards the stadium; storm had stronger updraft based on height of 50 dBZ echo	A/COA: Though flooding primary threat, needed to issue Severe Thunderstorm Warning for part of metro for large hail/damaging wind

Source	Cue	Assessment (A)/ Concern (C)/ Projection (P)/ Course of Action (COA)
AWIPS forecast model data and analysis of area sounding	Storms should move east at 10-15 mph	P: Expected the storm to move east slowly despite strong wind fields aloft A: Enhancing flood threat
Analysis of current observations and forecast model data	Atmosphere has not changed since earlier in the day	C: Knew atmosphere still holding a lot of water and a continued flash flood threat existed
ESPN	Ball game stopped for water	A: Verified that situation was as feared; this was some of our earliest ground truth for this flood
Phone call to another office	No reports of extreme rain in area the storm is passing through	P/A: Expected storms to produce hail, strong winds and flooding, was perplexed at the lack of verifying reports (although this is not uncommon in unpopulated areas)
AWIPS	Diffuse dry line	A: Not a classic day for tornadoes
Satellite picture	High cirrus clouds moving quickly to the northeast	A/P: (combined with other cues) May limit instability and make cap hard to break
Previous two days	Combination of previous two days with assessment of current day situation; numerical models forecast favorable severe storm conditions	P: Severe storms were likely in the next 24 hours
Manager's meeting	Discussed numerical model trends and SPC guidance	A/P: Agree with others that we wouldn't see significant weather events until evening and overnight
Internet weather page from own desk	Monitor trends in moisture, instability, radar, and cap strength	A/P: Cap was weakening more quickly than expected making storms more likely earlier than expected
Satellite and radar imagery from own desk	See whether storms were strengthening-looked for sign of boundary layer updrafts and increasing radar reflectivity	A/P: Storms were utilizing most unstable air in boundary layer and would be more severe

Source	Cue	Assessment (A)/ Concern (C)/ Projection (P)/ Course of Action (COA)
OK Mesonet, ARPS hourly analyses	Heating	C: Had doubts that enough heat would be present to initiate storms past cap
Balloon release	Layer of warmer air had developed	C: Began to question mental model of what might happen that day; became worried that something might happen earlier in the day
Satellite imagery	Dense clouds above were cutting off sunlight and thus heating	P: Wondered if they would still develop storms or if cirrus clouds would eliminate heating and no storms would develop
Balloon observation	Cap nearly breakable	P: Might mean that something severe might happen quickly if enough heating occurs
Satellite imagery from own desk	Storms “trying to pop up”	A: Atmosphere ripe for a severe storm P: Felt it was “going to go”
Satellite and radar	Several updraft regions/elevated showers evident	A: Cap broken locally
Satellite and radar	An updraft “took hold” and became a supercell	A: Previous conceptual model was wrong P: Realized this was going to be big and much sooner than expected
Storm spotter report by radio	Rotation of the cloud base within that supercell	A: Tornado formation increasingly likely
Radar	Another supercell storm developed behind the first	A: First storm not an isolated event P: Many storms may form
Radar	More cells continued to develop behind the others	A: May be looking at early stage of significant severe weather event
3-D reflectivity data	50 dBZ reflectivity at or above 30,000 ft	P: Storm will continue to grow and not dissipate, abundance of hail
Previous assessments; warnings issued	This was a “different” day; warnings issued were “standard”	C: Became concerned that these warnings would not get people’s attention
Radar, TV, spotter reports	Looked like a long-track, family-producing cell	C: Concerned that a tornado would develop and strike own home

Source	Cue	Assessment (A)/ Concern (C)/ Projection (P)/ Course of Action (COA)
Visual cue	Person who was at long-term forecasting desk was busy on the phone taking reports from highway patrol (of tornado sightings)	A: Believed they were shorthanded in that area COA: Thought should step in and help deal with the event
Visual cue	Person was able to come back to their desk	COA: Reassume oversight role
Radar, TV	Storm not turning away from direct path (still moving toward population center)	A/P: Catastrophe imminent
Reports from highway patrol	Small tornadoes with the storm	A: Optimum environment for tornadoes P: Tornadoes should persist
Helicopter video from the TV news	Tornado about 1 mile wide	P: Tornado would live for a long time and inflict severe damage; lives will be lost in potentially large numbers
Radar	Consistently strong reflectivity in tip of hook	A: Caused by suspended structural debris; catastrophe underway
TV coverage, previous storm-chasing experiences	Visual images of storm match images from storm-chasing experiences	A: A tornado of rare strength, size, and long life
AWIPS workstation graphical displays of satellite imagery	Indications of upper level jet streaks approaching area	A: Jet streaks enhance mesoscale upward vertical motion, helping to further destabilize the environment
Infra red image (stability with satellite overlaid) at AWIPS workstation	Areas of high instability	A: Increased risk of severe thunderstorms
Radar	Sharp horizontal differences in temperature or moisture	A: Tornadoes more likely to develop in vicinity of temperature/moisture boundaries
Science officer on previous shift	Not paying close attention to the radar when he came on the shift; gave the assessment as “no problem”	C: Wasn’t worried about severe weather
Management of previous shift	No one held over to work the next shift	P: No expectations for severe weather
Wind profilers and model soundings	Lot of shear in the environment	A: Good environment for supercell development
Wind profilers and model soundings	Winds had persistent rotation (high spin up)	A: Good environment for supercell development

Source	Cue	Assessment (A)/ Concern (C)/ Projection (P)/ Course of Action (COA)
88D radar storm-relative velocity products	Some rotation visible	COA: Need to switch and work the warning desk (radar)
88D radar storm-relative velocity products	Signature of velocity	A: Strong rotational signature suggestive of strong meso and potential for tornadic development
88D radar storm-relative velocity products	Persistent rotation for 2-3 volume scans (about 15 minutes)	A: Persistent rotation suggests mini-supercell, heightening risk for tornadic development
88D radar low-level reflectivity products	No “nice hooks”	C: Less cause for concern
Balloon reports	No great instability in the environment	A: Environment not “supportive” of severe storms
88D-radar low-level reflectivity products	Some subtle reflectivity	C: Cause for some concern
Wind profilers and model soundings	Modestly high shear	A: Upper air is turning, has helicity, rotation
88D-low-level storm-relative velocity products	Strong in/outward rotation (couplet)	C: Cause for concern
Earlier warnings	Other storms in the area with severe signatures	C: Cause for concern
National verification statistics	Pressure to limit false alarms	A: Factor for caution in issuing warning
Personal knowledge of population distributions in affected area	Population density lower in that area	A: Reason for caution in issuing warning
Other office	No call received	C: Decreased concern for severe weather
4 panel display-volume scans	Subtle signs that matched signs he’d seen before	A: Indicative of storms in the southeast
Other panes and products	Strong velocity signature	C: First cue of something suspicious
Other panes and products (looking for hooks and notches, comparing to velocity)	Strength of reflectivity not very impressive by itself	C: In conjunction with velocity, reflectivity sufficient cause for concern
88D reflectivity product; knowledge of population distributions	Storm had moved into a populated area	COA: Decision to warn
88D low-level storm-relative velocity product	Signature went away a little, then returned to the same intensity	A: Storm has persistence-indicative of mini-supercell; increased risk for tornado

HYPOTHETICAL CUES POSED TO THE SUBJECT-MATTER EXPERT

Source	Cue	Assessment (A)/ Concern (C)/ Projection (P)/ Course of Action (COA)
Spotter call	Report of a funnel or damage from a storm you didn't watch	A: Know you're not ahead of the storm
New information coming in	Mental model breaks down based on this new information	A: Might be paying too little attention to other cells

Note: The assessments, projections, concerns, and courses of action developed from these cues were specific to the context of a particular situation. The order of the cues as presented here is not meant to be chronological.

Appendix B - Interview Specific Cue Inventory

INTERVIEW SPECIFIC CUE INVENTORY

This is a sample cue inventory from one specific interview. Some cells do not contain information. We sent individual inventories to the appropriate SME and requested this additional information.

Source	Cue	Assessment (A)/ Concern (C)/ Projection (P)/ Course of Action (COA)
Own sense of smell (outside)	Smelled the Gulf of Mexico moisture in the air	
Visual cue outside	Noticed that the speed of the clouds was "different"	
Visual cue outside	Severity of downpour (necessitated a physical rescue of goldfish outside)	Atmosphere capable of producing huge amounts of water
Internet (from home)	Tornado watch issued for an area located west of his CWA, but moving east	
Internet radar (from home)	Strong vertical wind shear	Instability in atmosphere with potential for serious storms
Internet radar (from home)	Lack of supercells developing	Tornadoes not the threat that day
Radio and HAM radio (in car)	Nobody talking about the weather	Severe weather might surprise these people because they're not paying attention to it
	Storm moving slowly along the river basin	Storm could produce huge amounts of water; there's an increased chance for flash floods
	Reflectivity	
	Vertical structure	
	Line of thunderstorms approaching metro area	Not strong enough to be severe, wonders if they will be severe in the near future-
	Hailstorm moving towards the stadium	
		Expected the storm to move east
	Atmosphere has not changed since earlier in the day	Knew atmosphere still holding a lot of water
ESPN	Ball game stopped for water	
Phone call to another office	No reports of extreme rain in area the storm is passing through	